ELSEVIER Stochastic Processes and their Applications 51 (1994) 117–134

stochastic processes and their applications

# Conditional tail probabilities in continuous-time martingale LLN with application to parameter estimation in diffusions

David Levanony

*Department of Electrical Engineering, McGill University, 3480 University St., Montreal, Quebec, Canada H3A 2A7*

Received 16 July 1992; revised 10 May 1993

**Abstract**

Let $M$ be a continuous martingale, $h:\mathbb{R}_+ \to \mathbb{R}_+$ continuous and increasing such that $M(t)/h(\langle M\rangle_t) \to 0$ (a.s.) as $t \to \infty$. It is shown that w.p.l, large deviations type limits exist for a class of conditional probabilities which are induced on $(C([0, \infty), \|\cdot\|_\infty)$ by the tail processes $y^t(\cdot) = M(t+\cdot)/h(\langle M\rangle_{t+\cdot})$. This is obtained via a simple use of the Borell inequality for Gaussian processes, combined with a random time change argument. Results are applied to obtain convergence rates for the (conditional) tail probabilities of consistent parameter estimators in diffusion processes. This is followed by the derivation of efficient stopping rules. Finally, unconditional large deviations lower bounds for the tails of consistent estimators in diffusions are investigated via an extension of a well known direct method.

*Keywords:* Tail probabilities; Large deviations; Martingale LLN; Borell inequality; Parameter estimation; Diffusions.

## 1. Introduction

Central limit theorems (CLT) are probably the most widely known tools for the evaluation of convergence rates in various LLN (law of large numbers) statements with deterministic or random norming, see e.g. (Hall and Heyde, 1980). CLTs are similarly applied in parameter estimation as a measure of the convergence rate of consistent, off-line and recursive estimation schemes, e.g. the MLE (maximum likelihood estimator) (Hall and Heyde, 1980) and SA (stochastic approximation) (Kushner and Huang, 1979), respectively. These results, which are in the form

$$I_n^{1/2}(X_n-\theta) \xrightarrow{\mathrm{D}} \mathcal{N}(0,\Sigma), \quad X,\theta\in\mathbb{R}^d,\ \Sigma\in\mathbb{R}^{d\times d}, \quad I_n\uparrow\ \infty, \tag{1.1}$$


*SSDI* 0304-4149(93)E0047-I

are essentially of local type where the "almost" normal distribution (for large $n$'s) enables to compute the probabilities of the estimation error $X_n - \theta$ laying (at a fixed, large $n$) outside an $\mathbb{R}^d$ ball of order $I_n^{-1/2}$.

On the other hand, the LD (large deviations) approach is concerned with fixed (i.e. time independent) sets which, combined with the process under study, determine series of events which become "rare" as $n \to \infty$ in the sense that their probabilities decay exponentially fast. Such an approach is considered by Bahadur et al. (1980), Bahadur (1983), and Kester and Kallenberg (1986) who obtain lower bounds of the form

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\theta(|X_n - \theta| > \lambda) \geq -b(\lambda, \theta), \tag{1.2}$$

for consistent parameter estimators $\{X_n\}$ in various processes, in particular, i.i.d. processes (Bahadur et al., 1980; Kester and Kallenberg, 1986) and finite state, Markov chains (Bahadur, 1983) (where $b(\cdot,\cdot)$ is related to the Kullback–Leiber number). As was noted in (Bahadur et al., 1980), such a result cannot, in general, be extended to obtain upper bounds. Nevertheless, it is shown that the MLE in i.i.d. processes actually attains the lower bound in (1.2) (Bahadur et al., 1980; Kester and Kallenberg, 1986), a fact which makes it optimal in the appropriate sense. Note however that, as CLTs, such results are of local type (with some possible extensions, see Section 3.3).

Considerably stronger results are presented by Kushner (1984) and Dupuis and Kushner (1985) where LD lower *and* upper bounds for the tail probabilities of SA estimates are derived. More precisely, let $\{X_n\}$ be a sequence of consistent SA iterates and denote by $x^n(\cdot) \in C[0, T]$, $T < \infty$ a continuous time interpolation with $x^n(0) = X_n$. Then, for all sets $A \in (C[0, T], \|\cdot\|_T)$ (where $\|f\|_T \triangleq \sup_{0<s\leq T}|f(s)|$), it is shown in (Kushner, 1984; Dupuis and Kushner, 1985) that

$$\liminf_{n\to\infty} \lambda_n \log P(x^n \in A \mid x^n(0) = x) \geq -\inf_{\phi \in A^\circ, \phi(0) = x} S(\phi, T),$$

$$\limsup_{n\to\infty} \lambda_n \log P(x^n \in A \mid x^n(0) = x) \leq -\inf_{\phi \in \bar{A}, \phi(0) = x} S(\phi, T), \tag{1.3}$$

where $S$ is a "good" rate function (see (Dembo and Zeitouni, 1992; Deuschel and Stroock, 1989) and $\lambda_n \to 0$ depends on the SA gain sequence. It is important to note that such results are restricted to finite $T$'s. Infinite time horizons are considered by Dupuis and Kushner (1989) where *conditional* LD upper bounds (in the a.s. sense) are presented in the following form: Fix $\lambda > \delta > 0$, then, outside an $\omega$-null set it holds that

$$\limsup_{n\to\infty} \lambda_n \log P(|X_m - \theta| > \lambda, \text{ some } m \geq n \mid \mathscr{F}_n, |X_n - \theta| \leq \delta)$$

$$\leq -\inf_{T>0} \inf_{\substack{\phi \in C[0,T] \\ |\phi(0)| \leq \delta, \|\phi\|_T > \lambda}} S(\phi, T) < 0, \tag{1.4}$$

where $\{\mathscr{F}_n\}_{n\geq 1}$ is the underlying family of increasing $\sigma$-algebras.

Our main objective in this paper is to derive similar, infinite horizon bounds for continuous time parameter estimates. However, unlike Dupuis and Kushner (1989), we do not work directly with the estimation error process, but use an indirect approach. We first derive a class of general LD-type laws which characterize continuous-time, martingale LLN statements. This result is then applied to the parameter estimation problem. Such a framework is motivated by the fact that in many cases, various consistent estimators are closely related with (normalized) martingale processes. Consider for example a collection of real valued, linear-in-the-parameter diffusions which satisfy the following SDEs (stochastic differential equations),

$$\mathrm{d}x_t^\theta = \theta m(x_t^\theta, t)\,\mathrm{d}t + \mathrm{d}w_t, \quad x_0^\theta = 0, \tag{1.5}$$

where $m:\mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$ satisfies some regularity conditions and $\{w_t\}$ is a standard Brownian motion. Fix $\theta^* \in \mathbb{R}$ then, under rather weak conditions (Feigin, 1978; Levanony et al., 1993) the MLE $\{\hat{\theta}_t\}$ (based on $\{x_t^{\theta^*}\}$) is strongly consistent for $\theta^*$ and,

$$\hat{\theta}_t - \theta^* = \langle M\rangle_t^{-1} M(t) \quad \forall t > 0, \tag{1.6}$$

where $M(t) = \int_0^t m(x_s^{\theta^*}, s)\,\mathrm{d}w_s$ and $\langle M\rangle_t = \int_0^t m^2(x_s^{\theta^*}, s)\,\mathrm{d}s$ are a continuous-time martingale and its increasing process, respectively. Strong consistency is obtained by a martingale LLN (Revuz and Yor, 1991). A martingale CLT implies asymptotic normality in the form of (1.1) (Feigin, 1978; Levanony et al., 1993).

Now, due to (1.6), a LD law for tail probabilities of $\hat{\theta}_t - \theta^*$ can be actually derived from corresponding general martingale laws. Such an indirect approach can be extended to cases where, loosely speaking, (1.6) holds only asymptotically, i.e. when the parameter difference $\hat{\theta}_t - \theta^*$ is exponentially equivalent (Dembo and Zeitouni, 1992) to a normalized martingale, see Section 3.

Unfortunately, a law in the form of (1.4) for normalized martingale processes (such as in the RHS of (1.6)) cannot be obtained by an extension of finite horizon results (as has been done in (Dupuis and Kushner, 1989)). To illustrate this fact, let $W$ be a standard Brownian motion and consider, as a simplified example, the set of $C[0, \infty)$ processes $\{x^t\}$ where $x^t(s) = W(t+s)/(t+s)$. It can be shown that $\{x^t\}$ satisfy a Schilder-type LD principal on $C[0, \infty)$ equipped with a weak topology (such as the topology of uniform convergence on compacts, see another example in (Deuscheul and Stroock, 1989, Section 1.3)), or strong topology, $C[0, T]$ LD laws (i.e. on $(C[0, T], \|\cdot\|_T)$) where such infinite or finite horizon laws are characterized by the corresponding Schilder rate functions for $\sqrt{\varepsilon}W$ (Dembo and Zeitouni, 1992; Deuschel and Stroock, 1989). For finite horizon $[0, T]$ the rate function is $S(\phi, T) = \frac{1}{2}\int_0^T \dot{\phi}^2(s)\,\mathrm{d}s$, if $\phi$ is absolutely continuous (with $S(\phi, T) = +\infty$ otherwise). It is easy to see that an infinite horizon extension in the form of (1.4) leads to RHS $\equiv 0$ (take for example $\phi(s) = (\lambda/T)s$ to obtain $S(\phi, T) = \frac{1}{2}\lambda^2/T$ which results in $\inf_{T>0} S(\phi, T) = 0$).

Furthermore, even an attempt to prove a LD principal for $\{x^t\}$ directly on $(C[0, \infty), \|\cdot\|_\infty)$ (where $\|f\|_\infty \triangleq \sup_{s\geq 0}|f(s)|$), poses a major difficulty and may be impossible. This is due to the fact that exponential tightness, which is a key property in LD proofs (Dembo and Zeitouni, 1992; Deuschel and Stroock, 1989), fails to hold. A proof of exponential tightness involves showing that, roughly speaking, $x^t$ lies outside a compact set in $(C[0, \infty), \|\cdot\|_\infty)$ with a probability decaying exponentially fast (as $t \to \infty$). However, compacts in $\{f \in C[0, \infty) \mid \|f\|_\infty < \infty, \lim_{s \to \infty} f(s) = 0\}$ are characterized as being closed, bounded, equicontinuous and in addition, vanish *uniformly* at infinity (Munkers, 1975, p. 279). The last property fails to hold for $x^t$ (in the exponential sense above) and therefore rules out exponential tightness.

On the other hand, it turns out that, with a minor loss in generality, it is possible in our case (i.e. normalized martingale processes) to compute limits of the type of (1.4) by utilizing rather simple tools based on Borell's inequality for centered Gaussian processes (Adler, 1990) and random time change (Karatzas and Shreve, 1988; Revuz and Yor, 1991). The loss of generality is characterized by the fact that results are restricted to probabilities of events of the form $\{\|x^t\|_\infty > \lambda\}$, where, unlike standard LD theory, computations do not involve a general rate function (as $S$ above). Nevertheless, this loss of generality does not affect our main goal, namely, the evaluation of convergence rates, where events as the one above are precisely those we are interested in. Following Dupuis and Kushner (1989), we consider conditional probabilities where, contrary to (1.4), we obtain exact limits (i.e. the lower and upper bounds coincide). Those lead to rather simple stopping rules which may be very useful in practice. A completely different approach is used to derive *unconditional* LD lower bounds for a class of parameter estimation problems. This is done by an infinite horizon extension of the direct method of Bahadur et al. (1980). As one may intuitively expect, conditional LD laws are characterized by rates of decay which are faster than their unconditional counterpart, see example 3.8.

The paper is organized as follows: The next section is concerned with the martingale result, that is, the exponential rate of decay of conditional tail probabilities for normalized, continuous time martingales. An application to parameter estimation in diffusion processes is presented in Section 3. Exact limits of conditional tail probabilities for a class of consistent estimators are derived. This leads to simple and useful stopping rules. Finally, unconditional LD lower bounds are obtained via a direct approach taken from Bahadur et al. (1980).

## 2. A martingale conditional law

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space, $(M(t), \mathscr{F}_t)_{t \geq 0}$ a real valued, continuous local martingale with a quadratic variation process denoted by $\langle M \rangle$. Let $0 < h \in C[0, \infty)$ be nondecreasing and satisfy

**Condition A.** $\lim_{t \to \infty} \sqrt{t \log \log t}/h(t) = 0.$

Then, a martingale LIL (law of iterated logarithm) (Revuz and Yor, 1991) implies that

$$\lim_{t\to\infty} M(t)/h(\langle M\rangle_t) = 0 \text{ a.e. on } \{\langle M\rangle_t \uparrow \infty\}. \tag{2.1}$$

**Remark 2.1.** A sufficient condition for (2.1) is (cf. Revuz and Yor, 1991):

$$\int_\varepsilon^\infty h^{-2}(t)\,\mathrm{d}t < \infty \quad \forall\varepsilon > 0.$$

Define the $C[0, \infty)$ process

$$y^t(\cdot) = M(t + \cdot)/h(\langle M\rangle_{t+\cdot}). \tag{2.2}$$

We claim the following theorem.

**Theorem 2.2.** *Let $(M(t),\mathscr{F}_t)_{t\geq 0}$ be a continuous local martingale with $\langle M\rangle_t \uparrow \infty$ P-a.s. Then, under condition A, the processes $\{y^t\}$ satisfies*

$$\lim_{t\to\infty} \varphi_t \log P(\|y^t\|_\infty > \lambda \mid \mathscr{F}_t) = -\tfrac{1}{2}\lambda^2, \quad P\text{-a.s.}, \quad \forall\lambda > 0, \tag{2.3}$$

*where*

$$\varphi_t = \varphi_t(\langle M\rangle) = \sup_{s\geq 0} \frac{s}{h^2(\langle M\rangle_t + s)}.$$

The proof of Theorem 2.2 is based on the following Brownian law:

**Lemma 2.3.** *Let $(W(t),\mathscr{G}_t)_{t\geq 0}$ be a standard Brownian motion, $T_t \uparrow \infty$ a continuous, $\{\mathscr{G}_t\}$-stopping time process. Define $x^t(\cdot) = W(T_t + \cdot)/h(T_t + \cdot)$ where $0 < h \in C[0, \infty)$ is nondecreasing and satisfies condition A. Then*

$$\lim_{t\to\infty} \varphi_t \log P(\|x^t\|_\infty > \lambda \mid \mathscr{G}_{T_t}) = -\tfrac{1}{2}\lambda^2, \quad P\text{-a.s.}, \quad \forall\lambda > 0, \tag{2.4}$$

*with $\varphi_t = \varphi_t(T)$.*

**Proof of Theorem 2.2.** Let $(W(t),\mathscr{G}_t)_{t\geq 0}$ be a standard Brownian motion which is constructed via random time change of the martingale $(M(t),\mathscr{F}_t)_{t\geq 0}$, namely (Karatzas and Shreve, 1988; Revuz and Yor, 1991)

$$W(t) = M(\tau(t)), \quad \mathscr{G}_t = \mathscr{F}_{\tau(t)}, \quad \tau(t) = \inf\{s > 0 \mid \langle M\rangle_s > t\}. \tag{2.5}$$

It is well known that $\langle M\rangle_s$ is a $\{\mathscr{G}_t\}$-stopping time (Karatzas and Shreve, 1988; Revuz and Yor, 1991). Let $T_t = \langle M\rangle_t$, then $P(\cdot|\mathscr{F}_t) = P(\cdot|\mathscr{G}_{\langle M\rangle_t})$ $\forall t \geq 0$ which, together with a random time change argument (i.e. $M(t + s) = W(\langle M\rangle_{t+s})\,\forall s,t \in \mathbb{R}_+$,

$P$-a.s. (Karatzas and Shreve, 1988; Revuz and Yor, 1991)) imply that, outside an $\omega$-null set, independent of $\lambda$ or $t$

$$P(\|y^t\|_\infty > \lambda \mid \mathscr{F}_t) = P\left(\sup_{s \geq 0} \frac{|M(t+s)|}{h(\langle M \rangle_{t+s})} > \lambda \mid \mathscr{F}_t\right)$$
$$= P\left(\sup_{s \geq 0} \frac{|W(\langle M \rangle_{t+s})|}{h(\langle M \rangle_{t+s})} > \lambda \mid \mathscr{G}_{\langle M \rangle_t}\right)$$
$$= P\left(\sup_{r \geq 0} \frac{|W(\langle M \rangle_t + r)|}{h(\langle M \rangle_t + r)} > \lambda \mid \mathscr{G}_{\langle M \rangle_t}\right) = P(\|x^t\|_\infty > \lambda \mid \mathscr{G}_{T_t}),$$
$$\forall \lambda,\ t > 0.$$

(The change in the supremum in the next-to-last equality follows the fact that $h$, $W$ and $\langle M \rangle$ are continuous and that $\langle M \rangle_t \uparrow \infty$.) This together with (2.4), proves (2.3). □

**Remark 2.4.** Let $\Sigma_t = \sigma\{\langle M \rangle_t\}$ ( = the $\sigma$-field generated by $\langle M \rangle_t$). Then, by conditioning on $\Sigma_t$ (rather than on the larger $\sigma$-field $\mathscr{F}_t$) it can be shown that $\forall \lambda > \delta > 0$

$$\liminf_{t \to \infty} \varphi_t \log P(\|y^t\|_\infty > \lambda \mid \Sigma_t, |y^t(0)| \leq \delta) \geq -\tfrac{1}{2}\lambda^2, \quad P\text{-a.s.}, \tag{2.6}$$

$$\limsup_{t \to \infty} \varphi_t \log P(\|y^t\|_\infty > \lambda \mid \Sigma_t, |y^t(0)| \leq \delta) \leq -\tfrac{1}{2}(\lambda - \delta)^2, \quad P\text{-a.s.}, \tag{2.7}$$

The proof of (2.6)–(2.7) relies on corresponding Brownian bounds followed by a time change argument (exactly as the proof of Theorem 2.2). The only differences lie in some minor technicalities in the proof of the Brownian counterpart of (2.6)–(2.7). These involve preconditioning on $\mathscr{F}_t$ and smoothing. We omit the details.

**Remark 2.5.** Consider the case of Gaussian martingale $M$ i.e., when the increasing process $\langle M \rangle$ is deterministic. In this case, computation of unconditional probabilities results in

$$\lim_{t \to \infty} \bar{\varphi}_t \log P(\|y^t\|_\infty > \lambda) = -\tfrac{1}{2}\lambda^2, \qquad \bar{\varphi}_t = \sup_{s \geq 0} \frac{\langle M \rangle_t + s}{h^2(\langle M \rangle_t + s)}. \tag{2.8}$$

The proof of (2.8) is considerably simpler than its "random" counterpart (2.3) for obvious reasons. Note the basic difference between the rates of the"deterministic" and the "random" cases (i.e. (2.8) and (2.3), respectively) where in general, $\varphi_t < \bar{\varphi}_t\ \forall t > 0$. For example, in the case of $h(t) = t$, we have $\bar{\varphi}_t = 1/\langle M \rangle_t = 4\varphi_t$. This fact has some implications on the difference between the bounds of the conditional and unconditional LD laws, see Sections 3.3 and 3.4.

**Proof of Lemma 2.3.** Our aim is to show that the liminf and the limsup (as $t \to \infty$) of the term on the LHS of (2.4) are equally bounded (from below and from above, respectively). In order to simplify notation, we use throughout this proof the definition $P_t(\cdot) = P(\cdot \mid \mathscr{G}_{T_t})$ (with $E_t(\cdot) = E(\cdot|\mathscr{G}_{T_t})$).

*The upper bound.* The proof is based on the Borell inequality for centered Gaussian processes. The following brief summary relies on Adler (1990).

Let $X$ be a centered, real valued Gaussian process on some metric parameter space $T$. Assume that $X$ is separable and has a.s. bounded trajectories. Let $d(t, s) = \sqrt{E(X_t - X_s)^2}$ be the canonical (pseudo) metric which is induced by the process $X$ on the parameter space $T$ and assume that $T$ is totally bounded in $d$ (i.e. $\mathrm{diam}(T) = \sup_{t,s \in T} d(t, s) < \infty$). Define

$$\mu = E \sup_{t \in T} X_t, \quad \sigma^2 = \sup_{t \in T} E X_t^2. \tag{2.9}$$

Then by (Adler, 1990, Theorem 2.1)

$$P\left(\left|\sup_{t \in T} X_t - \mu\right| > \lambda\right) \leq 2 \exp -\tfrac{1}{2}\lambda^2/\sigma^2. \tag{2.10}$$

An easy consequence from (2.10) is that

$$P\left(\sup_{t \in T} |X_t| > \lambda\right) \leq 4 \exp -\tfrac{1}{2}(\lambda - \mu)^2/\sigma^2 \quad \forall \lambda > \mu. \tag{2.11}$$

The result is utilized below. Recall that $\{T_t\}$ is an a.s. continuous, $\{\mathscr{G}_t\}$-stopping time (with $T_t \uparrow \infty$). Hence, by the strong Markov property and time shift invariance of the Brownian motion, $B^t(\cdot) = W(T_t + \cdot) - W(T_t)$ is a standard Brownian motion, independent of $\mathscr{G}_{T_t}$. This makes $\beta^t(\cdot) = B^t(\cdot)/h(T_t + \cdot)$ a $P_t$-centered, continuous Gaussian process. In order to apply (2.11) to the process $\beta^t$ one first has to show that $\beta^t$ possesses $P_t$-a.s. bounded sample paths. In fact, we need that the $P_t$-a.s. boundedness holds for all large enough $t$'s, $P$-a.s. This is shown as follows: Let $\tau = \tau(\omega) = \inf\{t > 0 \mid T_t > 1\}$ then, $\tau < \infty$ *a.s.* and

$$\begin{aligned} \sup_{t \geq \tau} \|\beta^t\|_\infty &= \sup_{t \geq \tau} \sup_{s \geq 0} \frac{|W(T_t + s) - W(T_t)|}{h(T_t + s)} \\ &\leq 2 \sup_{t \geq \tau} \sup_{s \geq 0} \frac{|W(T_t + s)|}{h(T_t + s)} = 2 \sup_{r \geq 1} \frac{|W(r)|}{h(r)} < \infty, \quad P\text{-a.s.} \end{aligned} \tag{2.12}$$

where the last inequality is due to the a.s. continuous sample paths of $W/h$ on $[1, \infty)$, (2.1) and the fact that $T_\tau = 1$.

This obviously leads to

$$P_t(\|\beta^t\|_\infty < \infty) = 1 \quad \forall t \geq \tau, \quad P\text{-a.s.} \tag{2.13}$$

Furthermore, since $W(r)/h(r)$ satisfies (2.1) and $T_t \uparrow \infty$ one has

$$\limsup_{t \to \infty} \|\beta^t\|_\infty \leq 2 \lim_{t \to \infty} \sup_{r \geq T_t} |W(r)|/h(r) = 0, \quad P\text{-a.s.} \tag{2.14}$$

Now note that $\varphi_t = \varphi_t(T)$ is the "variance" which corresponds to $\beta^t$ (see (2.9)). By condition A and the fact that $T_t \uparrow \infty$, we obtain

$$\varphi_t = \sup_{s \geq 0} \frac{s}{h^2(T_t + s)} \leq \sup_{r \geq T_t} \frac{r}{h^2(r)} \to 0 \quad \text{as} \quad t \to \infty, \quad P\text{-a.s.} \tag{2.15}$$

Let $\mu_t = E_t \sup_{s \geq 0} \beta^t(s) = E_t \sup_{s \geq 0} B^t(s)/h(T_t + s)$. We claim that $\mu_t \to 0$ a.s. To prove this statement, recall that $\sup_{t \geq \tau} \|\beta^t\|_\infty$ is dominated by $Y \triangleq 2 \sup_{r \geq 1} |W(r)|/h(r)$. Now, since $\{W(r)/h(r),\ r \geq 1\}$ is an a.s. bounded, centered Gaussian process then, by (Adler, 1990, Lemma 3.1, Theorem 3.2), $EY < \infty$. Let $\mathscr{G}_\infty \triangleq \sigma\{\bigcup_{t \geq 0} \mathscr{G}_{T_t}\}$. Then, the integrability of $Y$, together with (2.14) enable to use a standard convergence theorem (e.g. Revuz and Yor, 1991, Corollary II. 2.4) to obtain

$$\lim_{t \to \infty} E(\|\beta^t\|_\infty | \mathscr{G}_{T_t}) = E\left( \lim_{t \to \infty} \|\beta^t\|_\infty | \mathscr{G}_\infty \right) = 0, \quad P\text{-a.s.} \tag{2.16}$$

The fact that by definition $0 < \mu_t \leq E(\|\beta^t\|_\infty | \mathscr{G}_{T_t})$ completes the proof of the claim.

Further, note that the initial condition $x^t(0)$ satisfies

$$\lim_{t \to \infty} x^t(0) = \lim_{t \to \infty} \frac{W(T_t)}{h(T_t)} = 0, \quad P\text{-a.s.} \tag{2.17}$$

which results due to (2.1). Finally, observe that

$$\sup_{t \geq \tau} \sup_{r,s \in \mathbb{R}_+} E_t(\beta^t(r) - \beta^t(s))^2 \leq 4 \sup_{t \geq \tau} \varphi_t < \infty, \quad P\text{-a.s.}$$

That is, the parameter space $\mathbb{R}_+$ is totally bounded in the (random) time dependent, canonical metrics which are induced by $\{\beta^t\}$ (see a definition before (2.9)), uniformly over $[\tau, \infty)$, a.s.

We are now ready to derive the upper bound: Choose $\varepsilon > 0$, $\delta \in (0, \lambda)$ and let $t_0 = t_0(\varepsilon, \delta) < \infty$ be s.t. $P(\sup_{t \geq t_0} (|x^t(0)| + \mu_t) > \delta) < \varepsilon$. Then, $\forall \lambda > \delta$, $t \geq t_0 \vee \tau$ we can write

$$\begin{aligned} \varphi_t \log P_t(\|x^t\|_\infty > \lambda) &= \varphi_t \log P_t\left( \sup_{s \geq 0} \left| \beta^t(s) + \frac{h(T_t)}{h(T_t + s)} x^t(0) \right| > \lambda \right) \\ &\leq \varphi_t \log P_t(\|\beta^t\|_\infty > \lambda - |x^t(0)|) \\ &\leq \varphi_t \log 4 \exp - \tfrac{1}{2}(\lambda - |x^t(0)| - \mu_t)^2/\varphi_t \xrightarrow[t \to \infty]{} -\tfrac{1}{2}\lambda^2, \\ &\qquad \text{w.p.} > 1 - \varepsilon, \end{aligned} \tag{2.18}$$

where the last inequality follows (2.11) and the fact that $x^t(0)$ and $\mu_t$ are $\mathscr{G}_{T_t}$ measurable. The limit on the RHS is obtained due to (2.15)–(2.17). Since (2.18) holds $\forall \varepsilon, \delta > 0$, $\lambda > \delta$ we may conclude that

$$\limsup_{t \to \infty} \varphi_t \log P_t(\|x^t\|_\infty > \lambda) \leq -\tfrac{1}{2}\lambda^2, \quad P\text{-a.s.}\ \forall \lambda > 0. \tag{2.19}$$

*The lower bound*: With a slight abuse of notation, let $\varphi_t(s) = E_t|\beta^t(s)|^2 = s/h^2(T_t + s)$ (with $\varphi_t = \sup_{s\geq 0} \varphi_t(s)$) and define the Gaussian function

$$\psi(\alpha) = \sqrt{\frac{2}{\pi}} \int_\alpha^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = P(|Z| > \alpha), \quad Z \sim \mathcal{N}(0,1).$$

Note that since under $P_t$, $\beta^t(s) \sim \mathcal{N}(0, \varphi_t(s))$, we have $P_t(|\beta^t(s)| > \alpha) = \psi(\alpha/\sqrt{\varphi_t(s)})$. Therefore,

$$\begin{aligned}
\varphi_t \log P_t(\|x^t\|_\infty > \lambda) &= \varphi_t \log P_t\left(\sup_{s\geq 0}|\beta^t(s) + \frac{h(T_t)}{h(T_t+s)}x^t(0)| > \lambda\right) \\
&\geq \varphi_t \log P_t\left(\sup_{s\geq 0}|\beta^t(s)| > \lambda + |x^t(0)|\right) \\
&\geq \varphi_t \log \sup_{s\geq 0} P_t(|\beta^t(s)| > \lambda + |x^t(0)|) \\
&= \varphi_t \log \sup_{s\geq 0} \psi\{(\lambda + |x^t(0)|)/\sqrt{\varphi_t(s)}\} \\
&= \varphi_t \log \psi\{(\lambda + |x^t(0)|)/\sqrt{\sup_{s\geq 0}\varphi_t(s)}\} \\
&= \varphi_t \log \psi\{(\lambda + |x^t(0)|)/\sqrt{\varphi_t}\} \\
&\geq \varphi_t \log \sqrt{\frac{2}{\pi}} \frac{(\lambda + |x^t(0)|)/\sqrt{\varphi_t}}{1 + (\lambda + x^t(0))^2/\varphi_t} \exp \\
&\quad - \tfrac{1}{2}(\lambda + |x^t(0)|)^2/\varphi_t \xrightarrow[t\to\infty]{} -\tfrac{1}{2}\lambda^2, \quad \text{a.s.}
\end{aligned} \tag{2.20}$$

The last inequality is a simple lower bound for $\psi$ (see (Karatzas and Shreve, 1988, p. 112). The limit is obtained due to the fact that $\varphi_t$ and $x^t(0) \to 0$ as $t \to \infty$. This, together with (2.20) result in (2.4). □

To end this section, consider for a moment $\mathbb{R}^d$ valued martingales. The following corollary is an easy consequence from Theorem 2.2.

**Corollary 2.6.** *Let $(M(t), \mathscr{F}_t)_{t\geq 0} = (M_i(t),\ 1 \leq i \leq d,\ \mathscr{F}_t)_{t\geq 0}$ be a continuous-time, vector valued martingale with $\langle M_i\rangle_t \uparrow \infty$, P-a.s. $\forall i = 1,2,\ldots,d$. Then, for any h as in Theorem* 2.2

$$\lim_{t\to\infty} \tilde{\varphi}_t \log P\left(\max_{1\leq i\leq d} \|y_i^t\|_\infty > \lambda \mid \mathscr{F}_t\right) = -\tfrac{1}{2}\lambda^2, \quad P\text{-a.s.} \tag{2.21}$$

*where* $y_i^t(s) = M_i(t+s)/h(\langle M_i\rangle_{t+s})$, $\tilde{\varphi}_t = \max_{1\leq i\leq d} \varphi_t(\langle M_i\rangle)$.

**Proof.** The probability in the LHS of (2.21) is bounded as follows

$$d \max_{1 \le i \le d} P(\|y_i^t\|_\infty > \lambda \mid \mathscr{F}_t) \ge P\left(\max_{1 \le i \le d} \|y_i^t\|_\infty > \lambda \mid \mathscr{F}_t\right)$$
$$\ge \max_{1 \le i \le d} P(\|y_i^t\|_\infty > \lambda \mid \mathscr{F}_t). \tag{2.22}$$

Now, by letting the terms in the proof of Lemma 2.3 depend on $i$ and taking the maximum over $1 \le i \le d$ (in the upper bound (2.19) and the lower bound (2.20), one concludes that indeed $\tilde{\varphi}_t$ is the norming sequence for the logarithms of both sides in (2.22). The fact that $\tilde{\varphi}_t \to 0$ and Theorem 2.2 (applied for both sides in (2.22)) complete the proof. □

**Remark 2.7.** Note that (2.22) holds also with the Euclidean norm replacing the max norm inside the probability on the LHS.

## 3. Parameter estimation in diffusion processes

In this section we examine the rate of convergence of the tails of consistent parameter estimators in diffusion processes. In particular, the martingale result (Theorem 2.2) is utilized in order to obtain exponential rates of decay for conditional probabilities of rare tail events of "first order efficient estimators" (Feigin, 1978; Hall and Heyde, 1980), among which, the MLE is shown to be optimal in the sense of having the fastest rate of convergence.

We end this section with a direct approach adopted from Bahadur et al. (1980) (see also (Bahadur, 1983; Kester and Kallenberg, 1986)) to the derivation of *unconditional* LD-type lower bounds for consistent parameter estimators. While in (Bahadur et al., 1980; Bahadur, 1983; Kester and Kallenberg, 1986) only local-type results are presented (for discrete i.i.d. sequences (Bahadur et al., 1980; Kester and Kellenberg, 1986) and finite state, discrete Markov chains (Bahadur, 1983)), we extend the general approach to infinite horizon, tail processes of consistent estimators in diffusions. Results are compared with the corresponding conditional LD law derived by the indirect, martingale approach.

### 3.1. Problem statement

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and consider a collection of real valued processes $\{X^\theta\}_{\theta \in \mathbb{R}}$ which satisfy the stochastic differential equations:

$$\mathrm{d}x_t^\theta = m(\theta, x_t^\theta, t)\,\mathrm{d}t + \sigma(x_t^\theta, t)\,\mathrm{d}w_t, \quad x_0^\theta = 0, \tag{3.1}$$

where $(w_t, \mathscr{F}_t)_{t \ge 0}$ is a standard Brownian motion, $\sigma^2$ is bounded from below and $m(\cdot,\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ satisfy some regularity conditions which ensure the existence and uniqueness of strong solutions to (3.1). Moreover, assume that the measures which are induced by $\{x_t^\theta, t \in [0, T]\}_{\theta \in \mathbb{R}}$ on $C[0, T]$ are mutually equivalent $\forall T < \infty$ (for specific conditions see e.g. (Borkar and Bagchi, 1982)).

Let $X = \{x_t,\ t \geq 0\}$ denote the observed process. Assume that there exists a $\theta^* \in \mathbb{R}$ s.t. $X = X^{\theta^*}$. Then, the log-likelihood is in the following form (Borkar and Bagchi, 1982; Feigin, 1978; Levanony et al., 1993)

$$L_t(\theta, X) = \int_0^t g(\theta, x_s, s)\,\mathrm{d}w_s - \tfrac{1}{2}\int_0^t g^2(\theta, x_s, s)\,\mathrm{d}s;$$

$$g(\theta, x, s) \triangleq \frac{m(\theta, x, s) - m(\theta^*, x, s)}{\sigma(x, s)}. \tag{3.2}$$

Assume that $L_t(\cdot, X)$ is twice differentiable $\forall t \in \mathbb{R}_+$, $P$-a.s. (see (Levanony et al., 1993) for sufficient conditions), denote its gradient by $U_t(\cdot, X)$ and define $U_t = U_t(\theta^*, X)$. It can be verified that $(U_t, \mathscr{F}_t)_{t \geq 0}$ is a continuous martingale. Denote by $I_t = \langle U \rangle_t$ its increasing process and assume throughout that $I_t \uparrow \infty$, $P$-a.s.

**Definition 3.1.** (Hall and Heyde, 1980). Let $\{\tilde{\theta}_t\}$ be an estimator of $\theta^*$. $\tilde{\theta}$ is called FOE (first order efficient) if there exists a (deterministic) constant $\gamma = \gamma(\theta^*)$ s.t.

$$I_t^{1/2}|\tilde{\theta}_t - \theta^* - \gamma I_t^{-1} U_t| \xrightarrow{\mathrm{P}} 0 \text{ as } t \to \infty\,. \tag{3.3}$$

It is well known that $\gamma \geq 1$ with $\gamma = 1$ for the MLE $\hat{\theta}$ (where $\hat{\theta}_t = \arg\sup_{\theta \in \mathbb{R}} L_t(\theta, X)$) see e.g. (Hall and Heyde, 1980, Theorem 6.1).

**Remark 3.2.** A martingale CLT implies that an FOE estimator $\tilde{\theta}$ with constant $\gamma$ satisfies (Hall and Heyde, 1980)

$$I_t^{1/2}(\tilde{\theta}_t - \theta^*) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \gamma^2), \tag{3.4}$$

which, by the discussion above, makes the MLE optimal in the sense of having the smallest asymptotic (weighted) variance.

### *3.2. Tail probabilities of consistent estimators: A martingale approach*

In the light of Theorem 2.2, one may expect that the martingale law (2.3) can be utilized to obtain a corresponding statement for any FOE estimator for which the convergence in (3.3) is sufficiently fast. More precisely, let $\{\tilde{\theta}_t\}$ be a continuous (a.s.) FOE estimator (with a constant $\gamma \geq 1$) and choose a non decreasing $h \in C[0, \infty)$ which satisfies Condition A. Define $v^t, y^t \in C[0, \infty)$ by

$$v^t(s) = \frac{I_{t+s}}{h(I_{t+s})}(\tilde{\theta}_{t+s} - \theta^*), \quad y^t(s) = U_{t+s}/h(I_{t+s}).$$

Then, in order to apply (2.3) to the (normalized) parameter difference one needs exponential equivalence between $\{v^t\}$ and $\{\gamma y^t\}$, namely that (Dembo and Zeitouni, 1992)

$$\lim_{t\to\infty} \varphi_t \log P(\|v^t - \gamma y^t\|_\infty > \varepsilon \mid \mathscr{F}_t) = -\infty, \quad P\text{-a.s.} \quad \forall \varepsilon > 0, \tag{3.5}$$

where $\varphi_t = \varphi_t(I) = \sup_{s\geq 0} s/h^2(I_t + s)$.

Note that (3.5) relies explicitly on $h$. Imposing on $h$ the following weak restriction

**Condition B.** $\exists b = b(h) \in (0, \infty)$ s.t.

$$\limsup_{t\to\infty} \frac{t}{h^2(t)} \inf_{s\geq 0} \frac{h^2(t+s)}{s} = b,$$

enables to consider a simpler condition (then (3.5)) which does not depend on $h$, is related explicitly to (3.3) and together with B, leads to the exponential equivalence (3.5):

**Condition C.**

$$\lim_{t\to\infty} a_t \log P\left( \sup_{r\geq t} I_r^{1/2} |\tilde{\theta}_r - \theta^* - \gamma I_t^{-1} U_t| > \varepsilon/\sqrt{a_t} \,|\, \mathscr{F}_t \right) = -\infty, \quad P\text{-a.s.}$$

for any $\{\mathscr{F}_t\}$ adapted process $\{a_t\}$, $0 < a_t \to 0$ and $\forall \varepsilon > 0$.

**Remark 3.3.** Condition C trivially holds for the MLE $\hat{\theta}$ in the cases where the measures induced by $\{X^\theta\}_{\theta\in\mathbb{R}}$ on $C[0, T]$ belong to a conditional exponential family (Feigin 1978; Hall and Heyde, 1980). In such cases, it holds that $\hat{\theta} - \theta^* = I^{-1} U$ which, in diffusion processes, are the linear-in-the-parameter models (1.5). Here, we only need that such a conditional exponential property may be reached asymptotically with a sufficiently fast exponential rate.

We can now state the following theorem.

**Theorem 3.4.** *Consider an FOE estimator $\tilde{\theta}$ with a constant $\gamma$ and let $h \in C[0, \infty)$, $h(t) \uparrow \infty$ satisfy conditions A, B. Assume that condition C holds. Then*

$$\lim_{t\to\infty} \varphi_t \log P\left( \sup_{r\geq t} \frac{I_r}{h(I_r)} |\tilde{\theta}_r - \theta^*| > \lambda \,|\, \mathscr{F}_t \right) = -\tfrac{1}{2}\lambda^2/\gamma^2, \quad P\text{-a.s.} \quad \forall \lambda > 0. \tag{3.6}$$

**Proof.** Since (2.3) holds for $\{y^t\}$, it suffices to show that $\{v^t\}$ and $\{\gamma y^t\}$ are exponentially equivalent (Dembo and Zeitouni, 1992), namely, that (3.5) holds. To this end, note that condition B, the definition of $\varphi_t$ in (3.5), together with the fact that $h(t)$ and $I_t$ are continuous and that $I_t \uparrow \infty$, implies that there exists a (deterministic) $\bar{b} = \bar{b}(h) \geq b$ s.t.

$$\sup_{\{t>0 \mid I_t \geq 1\}} \frac{I_t}{h^2(I_t)} \Big/ \varphi_t = \sup_{t\geq 1} \frac{t}{h^2(t)} \inf_{s\geq 0} \frac{h^2(t+s)}{s} = \bar{b} < \infty, \quad P\text{-a.s.} \tag{3.7}$$

This implies that, a.e. on $\{\omega \mid I_t \geq 1\}$

$$\sup_{r \geq t} \frac{I_r}{h^2(I_r)} = \sup_{s \geq 0} \frac{I_t + s}{h^2(I_t + s)} \leq \frac{I_t}{h^2(I_t)} + \sup_{s \geq 0} \frac{s}{h^2(I_t + s)} \leq (\bar{b} + 1)\varphi_t. \tag{3.8}$$

Hence, by combining this together with the definitions of $v^t$ and $y^t$ we have

$$\begin{aligned}
&\varphi_t \log P(\|v^t - \gamma y^t\|_\infty > \varepsilon \mid \mathscr{F}_t) \\
&= \varphi_t \log P\left( \sup_{r \geq t} \frac{I_r}{h(I_r)} \left| \tilde{\theta}_r - \theta^* - \gamma I_r^{-1} U_r \right| > \varepsilon \mid \mathscr{F}_t \right) \\
&\leq \varphi_t \log P\left( \sup_{r \geq t} \frac{I_r^{1/2}}{h(I_r)} \sup_{r \geq t} I_r^{1/2} \left| \tilde{\theta}_r - \theta^* - \gamma I_r^{-1} U_r \right| > \varepsilon \mid \mathscr{F}_t \right) \\
&\leq \varphi_t \log P\left( \sup_{r \geq t} I_r^{1/2} \left| \tilde{\theta}_r - \theta^* - \gamma I_r^{-1} U_r \right| > \frac{\varepsilon}{\sqrt{\bar{b} + 1}} \Big/ \sqrt{\varphi_t} \mid \mathscr{F}_t \right) \\
&\xrightarrow[t \to \infty]{} -\infty, \quad P\text{-a.s.}
\end{aligned}$$

where the last inequality follows (3.8) and the limit is obtained due to condition C.

Since this holds $\forall \varepsilon > 0$, the processes $\{v^t\}$ and $\{\gamma y^t\}$ are exponentially equivalent (Dembo and Zeitouni, 1992) which allows to infer (4.5) from the law for $\{y^t\}$ (i.e. (2.3) with $\lambda/\gamma$). □

In the following examples, only the MLE $\hat{\theta}$ is considered (Condition C is assumed to hold):

**Example 3.5.** $h(s) = s^{(1+\nu)/2}$, $\nu \in (0, 1]$: In this case $\varphi_t = 1/c(\nu) I_t^\nu$, $c(\nu) = (1 + \nu)^{1+\nu}/\nu^\nu$ and,

$$P\left( \sup_{r > t} I_r^{(1-\nu)/2} \left| \hat{\theta}_r - \theta^* \right| > \lambda \mid \mathscr{F}_t \right) \approx \exp - \frac{c(\nu)}{2} I_t^\nu \lambda^2.$$

In particular, for $\nu = 1$ one has $c(\nu) = 4$ and

$$P\left( \sup_{r \geq t} \left| \hat{\theta} - \theta^* \right| > \lambda \mid \mathscr{F}_t \right) \approx \exp - 2 I_t \lambda^2.$$

**Example 3.6.** $h(s) = \sqrt{s \log s}$ $(s > 1)$: It can be shown that Condition B holds with

$$b = \lim_{t \to \infty} \inf_{s \geq 0} \frac{(t + s)\log(t + s)}{s \log t} = 1,$$

which leads to

$$P\left( \sup_{r \geq t} \left( \frac{I_r}{\log I_r} \right)^{1/2} \left| \hat{\theta}_r - \theta^* \right| > \lambda \mid \mathscr{F}_t \right) \approx \exp - \tfrac{1}{2} \lambda^2 \log I_t = I_t^{-\lambda^2/2},$$

**Remark 3.7.** In the light of the CLT (4.4) (and Condition A) this example may be considered as a limit case.

Note that (3.6) may be utilized to derive stopping rules: For example in the classical case of $h(t) = t$, fix $\varepsilon > 0$, $0 < \delta < \lambda$ and let $T = T(\varepsilon, \delta, \lambda) < \infty$ be s.t. w.p. $> 1 - \varepsilon$

$$P\left(\sup_{r \geq t}\left|\hat{\theta}_t - \theta^*\right| > \lambda \mid \mathscr{F}_t\right) \leq \exp - 2I_t(\lambda - \delta)^2 \quad \forall t \geq T.$$

Define the stopping time

$$\tau(\eta, \beta) = \inf\{t > 0 \mid \exp - 2I_t\beta^2 < \eta\} = \inf\{t > 0 \mid I_t > \log \eta^{-1/2\beta^2}\}.$$

Then it holds that, w.p. $> 1 - \varepsilon$

$$P\left(\sup_{r \geq t} | \hat{\theta}_t - \theta^* | > \lambda \mid \mathscr{F}_t\right) \leq \eta \quad \forall t \geq T \vee \tau(\eta, \lambda - \delta). \tag{3.9}$$

**Example 3.8** (*The ergodic case*). Let Eqs. (3.1) be time homogeneous (i.e. the drift and diffusion functions $m$ and $\sigma$ do not depend on $t$ explicitly) and assume that $\forall \theta \in \mathbb{R}$ there exists an invariant measure $\mu_\theta$ s.t. for any Borel function $f: \mathbb{R}^2 \to \mathbb{R}$, $\theta'$, $\theta \in \mathbb{R}$

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t f(\theta', x_s^\theta)\, \mathrm{d}s = \int_{\mathbb{R}} f(\theta', x)\, \mathrm{d}\mu_\theta(x) \triangleq E_\theta f(\theta', x). \tag{3.10}$$

Consider $h(t) = t$. Then $\varphi_t = 1/4I_t = 1/4 \int_0^t m_\theta^2(\theta^*, x_s)\, \mathrm{d}s$ and Theorem 3.4 together with (3.10) lead to

$$\lim_{t \to \infty} \frac{1}{t} \log P\left(\sup_{r \geq t}\left|\tilde{\theta}_r - \theta^*\right| > \lambda | \mathscr{F}_t\right) = -2\lambda^2 E_{\theta^*} m_\theta^2(\theta^*, x)/\gamma^2, \quad \forall \lambda > 0. \tag{3.11}$$

It is demonstrated below, that, in a large class of problems, for any estimator with $1 \leq \gamma < 2$ (e.g. the MLE), the RHS of (3.11) is larger (in absolute value) than the corresponding bound in the unconditional LD law (for all sufficiently small $\lambda$'s). This implies that the conditional probabilities decay faster than the unconditional ones. This may be expected while noting the inherent difference between the normalizing sequences $\varphi_t$, and $\bar{\varphi}_t$ (see definition in (2.8)) of conditional and unconditional LD laws, respectively (Remark 2.5). In he case above, the fact that $\varphi_t = \frac{1}{4}\bar{\varphi}_t$ suggests that, for the MLE (i.e. $\gamma = 1$), the factor 2 on the RHS could probably be replaced by a factor of 1/2. This conjecture is shown to hold (asymptotically as $\lambda \to 0$). Results follow an extension of the direct approach which is due to Bahadur et al. (1980).

### *3.3. Unconditional, LD-type lower bounds: A direct approach*

Fix $\theta$ and let $\tilde{\theta}_t(X^\theta)$ be an estimator which is based on the observation of $\{x_s^\theta, 0 \leq s \leq t\}$, rather than on the paths of $X = X^{\theta^*}$ in $[0, t]$ (we denote, as before, $\tilde{\theta}_t = \tilde{\theta}_t(X)$).

**Definition 3.9.** $\tilde{\theta}$ is a uniformly strong consistent estimator if

$$\tilde{\theta}_t(X^\theta) \to \theta, \quad \text{as } t \to \infty, \quad P\text{-a.s.} \quad \forall \theta \in \mathbb{R}.$$

Recall that $g(\theta, \cdot, \cdot) = [m(\theta, \cdot, \cdot) - m(\theta^*, \cdot, \cdot)]/\sigma(\cdot, \cdot)$ and define the martingale $G(\theta, X^\theta)$ and its increasing process $\langle G(\theta, X^\theta) \rangle$

$$G_t(\theta, X^\theta) = \int_0^t g(\theta, x_s^\theta, s)\,\mathrm{d}w_s, \quad \langle G(\theta, X^\theta) \rangle_t = \int_0^t g^2(\theta, x_s^\theta, s)\,\mathrm{d}s.$$

Then it can be verified that

$$L_t(\theta, X^\theta) = \log \frac{\mathrm{d}v_t^\theta}{\mathrm{d}v_t^{\theta^*}}(X^\theta) = G_t(\theta, X^\theta) + \tfrac{1}{2}\langle G(\theta, X^\theta) \rangle_t, \tag{3.12}$$

where $v_t^\theta$, $v_t^{\theta^*}$ are the measures which are induced by $X^\theta$ and $X^{\theta^*}$(respectively) on the space of continuous functions on $[0, t]$.

The following condition is considered below.

**Condition D.** There exists a deterministic $\{\alpha_t\}$, $0 < \alpha_t \to 0$ s.t.

$$\liminf_{t \to \infty} \alpha_t \langle G(\theta, X^\theta) \rangle_t > 0, \quad P\text{-a.s.} \quad \forall \theta \neq \theta^*,$$

$$\limsup_{t \to \infty} \alpha_t \langle G(\theta, X^\theta) \rangle_{t+T} < \infty, \quad P\text{-a.s.} \quad \forall \theta, \quad T < \infty.$$

Define

$$K(\theta, T) = \inf\{a \in \mathbb{R} \mid \lim_{t \to \infty} P(\alpha_t L_{t+T}(\theta, X^\theta) \leq a) = 1\}. \tag{3.13}$$

Note that condition D above, Eq. (3.12) and a martingale LLN (i.e. $G_t(\theta, X^\theta)/\langle G(\theta, X^\theta) \rangle_t \to 0$, a.s.) imply that

$$\liminf_{t \to \infty} \alpha_t L_{t+T}(\theta, X^\theta) = \tfrac{1}{2} \liminf_{t \to \infty} \alpha_t \langle G(\theta, X^\theta) \rangle_{t+T} > 0, \ P\text{-a.s.} \ \forall T < \infty, \ \theta \neq \theta^*. \tag{3.14}$$

This enables to redefine $K(\theta, T)$ as

$$K(\theta, T) = \inf\{a > 0 \mid \lim_{t \to \infty} P(\tfrac{1}{2}\alpha_t \langle G(\theta, X^\theta) \rangle_{t+T} \leq a) = 1\}. \tag{3.15}$$

The following lemma which is due to Bahadur et al. (1980) is reformulated to serve our purpose:

**Lemma 3.10** (Bahadur et al., 1980, Theorem 2.1). *Assume that condition D holds and let* $\{A_t^T(X^\theta)\}_{t \geq 0}$ *be a sequence of* $\mathscr{F}_{t+T}$ *-measurable events s.t.*

$$\liminf_{t \to \infty} P(A_t^T(X^\theta)) > 0, \quad \forall T \in [0, \infty). \tag{3.16}$$

*Then*

$$\liminf_{t\to\infty} \alpha_t \log P(A_t^T(X)) \geq -K(\theta, T), \quad \forall T \in [0, \infty). \tag{3.17}$$

**Remark 3.11.** The proof relies on the absolute continuity of the measures induced by $X^\theta$ on $C[0, t+T]$ with respect to the corresponding measures induced by $X$. This obviously rules out the possibility to consider *directly* the case $T = \infty$ (which is the one we are interested in). Nevertheless, as is shown below, lower bounds for infinite horizon, tail probabilities for consistent estimators are a rather straightforward conclusion from finite horizon results.

**Theorem 3.12.** *Let conditions D hold and assume that $\tilde{\theta}$ is a uniformly strong consistent estimator. Then*

$$\liminf_{t\to\infty} \alpha_t \log P(\sup_{r\geq t} |\tilde{\theta}_r - \theta^*| > \lambda) \geq - \inf_{|\theta-\theta^*|>\lambda} K(\theta, 0) \geq - \inf_{|\theta-\theta^*|>\lambda} \bar{K}(\theta), \tag{3.18}$$

*where*

$$\bar{K}(\theta) = \tfrac{1}{2} \operatorname*{ess\,sup}_{\omega} \limsup_{t\to\infty} \alpha_t < G(\theta, X^\theta) >_t. \tag{3.19}$$

**Proof.** Fix $\lambda > 0$ and choose $\theta$ s.t. $|\theta - \theta^*| > \lambda$. Define

$$A_t^T(X^\theta, \lambda) = \{ \sup_{s\in[0,T]} |\tilde{\theta}_{t+s}(X^\theta) - \theta^*| > \lambda \}.$$

Recall that $\tilde{\theta}(X^\theta)$ is strongly consistent for $\theta$ hence, by the choice of $\theta$,

$$\lim_{t\to\infty} P(A_t^T(X^\theta, \lambda)) = 1, \tag{3.20}$$

which, by Lemma 3.10 implies that

$$\begin{aligned}\liminf_{t\to\infty} \alpha_t \log P(A_t^T(X, \lambda)) &= \liminf_{t\to\infty} \alpha_t \log P(\sup_{s\in[0,T]} |\tilde{\theta}_{t+s} - \theta^*| > \lambda) \\ &\geq -K(\theta, T). \end{aligned} \tag{3.21}$$

Since this holds $\forall\theta$, $|\theta - \theta^*| > \lambda$ we obviously have

$$\liminf_{t\to\infty} \alpha_t \log P(\sup_{s\in[0,T]} |\tilde{\theta}_{t+s} - \theta^*| > \lambda) \geq - \inf_{|\theta-\theta^*|>\lambda} K(\theta, T). \tag{3.22}$$

Furthermore,

$$\begin{aligned}&\liminf_{t\to\infty} \alpha_t \log P(\sup_{s\geq 0} |\tilde{\theta}_{t+s} - \theta^*| > \lambda) \\ &\quad = \liminf_{t\to\infty} \alpha_t \sup_{T>0} \log P(\sup_{s\in[0,T]} |\tilde{\theta}_{t+s} - \theta^*| > \lambda)\end{aligned}$$

$$\geq \sup_{T>0} \liminf_{t\to\infty} \alpha_t \log P\big(\sup_{s\in[0,T]} |\tilde{\theta}_{t+s} - \theta^*| > \lambda\big)$$
$$\geq - \inf_{|\theta-\theta^*|>\lambda} \inf_{T>0} K(\theta, T), \tag{3.23}$$

where the last inequality is due to (3.22). Now, due to (3.15) we can write that

$$\inf_{T>0} K(\theta, T) = \inf_{T>0} \inf\left\{a > 0 \,|\, \lim_{t\to\infty} P(\tfrac{1}{2}\alpha_t < G(\theta, X^\theta) >_{t+T} \leq a) = 1\right\}$$
$$\leq \inf\left\{a > 0 \,|\, P\left(\tfrac{1}{2}\limsup_{t\to\infty} \alpha_t < G(\theta, X^\theta) >_t \leq a\right) = 1\right\}$$
$$= \tfrac{1}{2} \operatorname*{ess\,sup}_{\omega} \limsup_{t\to\infty} \alpha_t \langle G(\theta, X^\theta)\rangle_t.$$

The third and last inequality completes the proof. □

We end with an application of Theorem 3.12 in the ergodic case:

**Example 3.8** (*continued*). Let (3.10) hold and take $\alpha_t = 1/t$. Then

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t g^2(\theta, x_s^\theta)\,ds = E_\theta g^2(\theta, x) = K(\theta, 0) = \bar{K}(\theta),$$

which by Theorem 3.12 leads to

$$\liminf_{t\to\infty} \frac{1}{t} \log P(\sup_{r\geq t} |\tilde{\theta}_r - \theta^*| > \lambda) \geq -\tfrac{1}{2} \inf_{|\theta-\theta^*|>\lambda} E_\theta g^2(\theta, x), \tag{3.24}$$

for any uniformly consistent estimator $\tilde{\theta}$. Note that for small $\lambda$ one has

$$E_\theta g^2(\theta, x) = |\theta - \theta^*|^2 E_\theta m_\theta^2(\theta^*, x) + o(|\theta - \theta^*|^2).$$

Hence, since $E_\theta m_\theta^2(\theta^*, x) \to E_{\theta^*} m_\theta^2(\theta^*, x)$ as $\theta \to \theta^*$, we obtain

$$\lim_{\lambda\downarrow 0} \liminf_{t\to\infty} \frac{1}{\lambda^2 t} \log P(\sup_{r\geq t} |\tilde{\theta}_r - \theta^*| > \lambda) \geq -\tfrac{1}{2} E_{\theta^*} m_\theta^2(\theta^*, x). \tag{3.25}$$

Compared with (3.11), this implies that (at least in the ergodic case) the unconditional probabilities decay slower than their conditional counterparts for small $\lambda$'s (and $\gamma < 2$). In the case of the MLE, the exponential rate in the conditional law is, as expected, higher than the one in unconditional law (3.25) by a factor of 4.

**Remark 3.13.** While Theorem 3.12 holds in the case of vector valued parameters (with $|\cdot|$ denoting the Euclidean norm), the conditional law (3.6) cannot be extended directly to the $\mathbb{R}^d$ parameter case. Recall that (3.6) is derived from the martingale law (2.4) and relies on the fact that (1.6) holds (asymptotically). In the $\mathbb{R}^d$ counterpart of (1.6) $\langle M\rangle$ is a $d \times d$ matrix, a fact which prevents a direct use of the vector valued martingale law (2.22). It seems however, that it is possible to derive bounds (rather than exact limits) for the $\mathbb{R}^d$ valued parameter case, based on (2.22). This point requires further study.

## References

R.J. Adler, An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes, Lecture Notes: Monograph Series 12 (Institute of Mathematical Statistics, 1990).

R.R. Bahadur, S.L. Zabell and J.C. Gupta, Large deviation, tests and estimates, in: I.M. Chakravarti, ed., Asymptotic Theory of Statistical Tests (Academic Press, New York, 1980) 33–64.

R.R. Bahadur, Large deviations of the maximum likelihood estimate in the Markov chain case, in: M.H. Rizvi, J.S. Rostag and D. Siegmund, Eds., Recent Advances in Statistics (Academic Press, New York 1983) 273–286.

V. Borkar and A. Bagchi, Parameter Estimation in continuous-time stochastic processes, Stochastics 8 (1982) 193–212.

A. Dembo and O. Zeitouni, Large Deviations and Applications (Jones and Bartlett, Boston, MA, 1992).

J.D. Deuschel and D.W. Stroock, Large Deviations (Academic Press, New York, 1989).

P. Dupuis and H.J. Kushner, Stochastic approximations via large deviations: asymptotic properties, SIAM J. Control Optim: 23 (1985) 675–696.

P. Dupuis and H.J. Kushner, Stochastic approximation and large deviations: upper bounds and w.p.1 convergence, SIAM J. control Optim 27 (1989) 1108–1135.

P.D. Feigin, Maximum likelihood estimation for continuous time stochastic processes, Adv. in Appl. Probab. 8 (1978) 712–736.

P. Hall and C.C. Heyde, Martingale Limit Theory and Its Application (Academic Press, New York, 1980).

I. Karatzas and S. Shreve, Brownian Motion and Stochastic Calculus (Springer, Berlin, 1988).

A.D.M. Kester and W.C.M. Kallenberg, Large deviations of estimators, Ann. Statist. 14 (1986) 648–664.

H.J. Kushner and H. Huang, Rates of convergence for stochastic approximation type algorithms, SIAM J Control Optim 17 (1979) 607–617.

H.J. Kushner, Asymptotic behavior of stochastic approximation and large deviations, IEEE Trans. Automatic Control AC-29 (1984) 984–990.

D. Levanony, A. Shwartz and O. Zeitouni, Recursive identification in continuous-time stochastic processes, Stochastic Process Appl. 49 (1994) 245–275.

J.R. Munkers, Topology: A First Course (Prentice-Hall, Englewood Cliffs, NJ, 1975).

D. Revuz and M. Yor, continuous Martingales and Brownian Motion (Springer, Berlin 1991).